2021

Exploring the minimum number of trials needed to accurately detect concealed information using EEG

Graham, R.

Graham, R. (2021) 'Exploring the minimum number of trials needed to accurately detect concealed information using EEG', The Plymouth Student Scientist, 14(2), pp. 532-547.
http://hdl.handle.net/10026.1/18514

The Plymouth Student Scientist
University of Plymouth

*The Plymouth Student Scientist, 2021, **14,** (2), 532-547*

Exploring the minimum number of trials needed to accurately detect concealed information using EEG

Rhian Graham

Project Advisor: Giorgio Ganis School of Psychology, University of Plymouth, Drake Circus, Plymouth, PL4 8AA

Abstract

Deception can have serious consequences, therefore detection methods such as the Electroencephalogram (EEG) are vital. However, they can be perceived as time-consuming thus causing limited usage. The purpose of this study was to discover the minimum number of trials needed to accurately detect deception. It was hypothesised that accuracy would increase with number of trials (due to reduced effects of noise) but would then fall (due to subject's fatigue). Due to the Coronavirus, this study was restricted to 9 participants within our 'Lab Bubble' of third-year Psychology students at the University of Plymouth. A Concealed Information Test (CIT) was conducted, using target stimuli (a random date requiring a button press response), probe stimuli (subjects' date of birth to conceal), and irrelevant stimuli (random dates). Participants wore a wireless EEG cap with 7 electrodes which recorded their brain activity and P300 responses. The results revealed that participants consistently produced higher amplitude responses to probe than to irrelevant stimuli, highlighting the accuracy of this method. At 10 trials deception was hard to detect, as noise caused variability in probe stimuli P300 responses. The accuracy dramatically increased until 40 trials and peaked at 60 trials. The minimum number of trials recommended was 50 as it was highly accurate and less time-consuming. Demonstrating the efficiency of the EEG (by discovering the minimum number of trials required) may encourage wider use. This has salient implications in law enforcement and security clearances, as it has the potential to prevent crimes, thus protecting both individuals and society.

Keywords: EEG, deception, neuroscience, brainwaves, P300.

Introduction

Deception is evident within over 20% of social interactions (DePaulo, 1996). It is defined as the act of deliberately causing someone to have false beliefs (Carlson, 2010). Masip et al. (2004) expanded on this definition by adding that it is still deception even if the fabrication, manipulation or concealment of the emotional or factual information is unsuccessful. The concealment of information is defined as deception when there is a professional obligation and clear expectation for the person to provide information (Carlson, 2010). Deception is especially hard to detect when communicated under a conflict of interest (Grazioli, 1998). Whilst deception can sometimes prove useful, it can also lead to negative outcomes, for example wars and unsuccessful arrests. King and Dunn (2010) described the significance of deception in law enforcement, more specifically, in suspect interviews and crime prevention whereby criminals may conceal information to evade arrest. In addition to this, embedded lies which only change some details of the truth (Vrij, 2008) were found to be commonly used by criminal suspects (Hartwig et al., 2007). Also, Vrij (2008) found that self-orientated lies, ones given to avoid punishment, were common in adults. Bussey (1992) replicated this finding in children and discovered that, from the age of 4, children can deliberately lie with the motive of escaping punishment. Therefore, many types of lies can be involved in deception and the concealment of information, which is particularly detrimental in forensics. This highlights the need for a reliable method to detect such deception.

A Concealed Information Test (CIT) can be used, which Lykken (1959) first introduced as the Guilty Knowledge Test. Rosenfeld (2011) described a 3 stimulus CIT which incorporates probes (relevant and infrequently presented stimuli), irrelevants (other stimuli from the same category that are frequently presented), and targets (irrelevant stimuli that require a response). Subjects respond differently to probes, thus indicating their guilt. This method revealed a 93.9% correct classification of guilt (Lykken, 1959). The CIT was initially used with polygraphs measuring the peripheral physiological response of skin conductance rates (SCR). Selle et al. (2016) explained the Orientating theory whereby participants have a heightened SCR response to relevant items (probes). The orientating response is enhanced when the stimulus is significant to the subject, causing a bigger SCR (Sokolov, 1963). This method is prevalent in Japan as it is used 5000 times a year and has been admissible in court since the 1960s (Osugi, 2011). However, Vrij (2008) discredited the use of polygraphs because incorrect feedback prevents the discovery of errors. Additionally, Rosenfeld (2011) questioned the use of the Autonomic Nervous System and skin conductance tests as they only reveal indirect responses to stimuli, whereas using an EEG would ensure direct brainwave responses to stimuli.

One alternative to polygraphs is using the CIT in Functional Magnetic Resonance Imaging (fMRI) which measures changes in blood flow and oxygenation. A limitation of this is that it can be uncomfortable for subjects, due to restricted movement and confined space (Vrij, 2008). Therefore, any subjects with claustrophobia cannot participate, thus limiting the applicability of this method. Similarly, Ganis et al. (2011), questioned the utility of fMRI due to covert countermeasures. Countermeasures are techniques used by subjects on control items to hide any response to the significant items, they can be physical (biting tongue) or mental (counting backwards), and are difficult to detect (Honts & Kircher, 1994). Ganis et al. (2011) found 100% accuracy of fMRI without countermeasures. The subjects were then trained to use countermeasures, to reflect criminals' systematic use of them. This resulted in a drop to 33% accuracy, reflecting the vulnerability of fMRI to countermeasures. Thus, the method proved to be unreliable due to the prevalence of countermeasures in forensic settings. Other limitations of using fMRI are the expense and length of time taken to carry out the experiment (Vrij, 2008). Luck (2014) stated that an EEG is a better alternative to fMRI as it is less expensive.

Another alternative is using the CIT to detect deception through an EEG. The EEG uses electrodes placed onto the subject's scalp to measure the electrical potential (voltage) and record the waveforms to show their brain activity (Luck, 2014). The probe stimuli are meaningful to subjects who are concealing information, which causes the elicitation of a P300 response (Johnson, 1986). This is an Event Related Potential (ERP) whereby the stimuli elicits a positive deflection in brainwaves (Fabiani, 1987). It peaks between 300-800ms after the event, revealing recognition of information (Verschuere et al., 2011). Rosenfeld (1998) found that the target stimuli also elicits the P300 because the response task causes the subjects to attend to it. A disadvantage of using EEG is that the voltage recording incorporates brain activity signals and biological background noise, the latter dominating the signal-to-noise ratio thus reducing the clarity of the results (Scheer et al., 2005). Therefore, the number of trials performed with each subject is significant, as too few would lead to too much noise and less reliable results and too many could cause fatigue in the subjects. To counter this, the average ERP is calculated over multiple trials, thus increasing the accuracy of results (Luck, 2014). To further limit the background noise, this study reduced movement artefacts through use of verbal instructions to participants.

Within the CIT in EEG probes of varying types and modalities can be used. The stimuli can be presented in visual or auditory modalities. Misaka et al. (2009) conducted a within-subjects study whereby 14 undergraduates were exposed to probes verbalised in a digitalised human voice and written on a computer screen. Both conditions displayed highly accurate information processing. However, the visual stimuli elicited a larger P300, allowing for easier detection of deception, although the auditory stimuli did maintain arousal. Thus, this study used visual stimuli and incorporated a button press task to ensure arousal and attention were maintained. Another way the probes can vary is by type. Ganis and Schendan (2013) conducted a 3 stimulus CIT consisting of semantic autobiographical (subjects' date of birth) and episodic autobiographical (a date learnt at beginning of experiment) probes. Due to the significance to the subject and the sensitivity of the P300 to knowledge and semantic systems in the neocortex, the semantic probes produced larger P300, thus increasing the accuracy of deception detection. Therefore, this experiment used semantic information (subjects' date of birth) as probes and episodic information (a random date) for the target stimuli. This was to ensure a more prominent P300 response was elicited by the probes.

Detecting deception using the P300 is of high accuracy and utility. Lefebvre et al. (2007) conducted a facial recognition study with time delays between the crime and the line-up. The results revealed the P300 was a reliable indicator of recognition with all time delays. Even when the subjects' accuracy decreased after a week, the P300 effect was still strong, highlighting its utility. Also, when the culprit was absent from the line-up there was no P300 effect, thus revealing the efficiency of this technique. A time delay between a crime and the police interview is inevitable, therefore it is necessary for the P300 to not be affected by prolonged retention intervals. However, Lefebvre et al. (2007) studied groups so, to further test the reliability of the P300, this experiment investigated the effects on individual cases. Furthermore, the present study used the P300 as an indicator of familiarity, but using written text as opposed to faces. As well as this, there was no time delay between the event and testing due to the nature of the stimuli (a semantic autobiographical probe), therefore the risk of subjects' accuracy decreasing was inhibited.

The efficiency of using an EEG is debated, this issue is addressed in this experiment. Luck (2014) argued that this method is efficient as it provides a record of ongoing brain activity in real time. Furthermore, Rosenfeld (2011) found it to be 85-100% accurate as meaningful information produced larger P300s. However, there is mixed evidence of the applicability of this method. Luck (2014) stated that this covert measure of deception is beneficial as it can be used on subjects who struggle with behavioural responses (such as young children or those with neurological deficits). In contrast, Rosenfeld (2011) argued that, due to the effort required to learn about and use an EEG, law enforcement agencies could be less inclined to adopt this technique. Therefore, this study expanded on previous experiments by testing the minimum number of trials needed for an EEG to accurately detect deception using the CIT, thus improving its applicability. Law enforcement agencies could then be more inclined to use this method as it would require fewer generations of trials, and therefore, be less time consuming. In contrast, Boudewyn et al (2018) argued that the minimum number of trials is dependent upon the design of the study (the sample size, effect magnitude and noise level). Nevertheless, finding the minimum would help to standardise EEG techniques, thus increasing their validity. However, Ganis and Schendan (2013) stated that having more trials improved the noise-to-signal ratio. In contrast, Lynn (2013) found that repeated presentation of the stimuli decreased novelty, causing the orientation response to habituate. This asserts that too many trials would decrease the subjects' response to meaningful stimuli, making deception harder to detect, thus reiterating the importance of finding the minimum number of trials required.

This experiment used CIT in EEG, due to the limitations of the other techniques (e.g. polygraphs and fMRI). Semantic autobiographical visual stimuli (of the subjects' date of birth written on computer screen) were used, as this type of probe elicits a stronger P300, making deception easier to detect. Furthermore, the minimum number of trials needed to detect concealed information was tested in order to reduce the effects of background noise and increase the applicability of this method. It was predicted that the P300 would be elicited when subjects recognise the stimuli and that an inverted-U result would be found. Yerkes and Dodson (1908) described an inverted-U curve of performance, whereby performance is poor then increases to optimum before falling again. The results would be poor initially due to small ERP signals compared to noise, then would rise to optimum, before falling as the subjects become more fatigued or the orientation effect habituates.

Methodology
Participants
Nine undergraduates (1 male and 8 females, aged 20-22 years old) were recruited from PSYC605 (a third-year Psychological Research Project module) at the University of Plymouth. They formed a 'Lab Bubble' and were tested at different times to minimise exposure due to Coronavirus precautions. Participants were mostly right-handed (with 1 left-handed) and all had normal or corrected-to-normal vision. Eight participants spoke English as their first language, and one was bilingual (English/Mandarin). Prior to the study, subjects reported their date of birth and were shown a list of dates and asked if any were of personal significance; these were removed. To test their eligibility to participate, all subjects completed a Health Screening Questionnaire (see Appendix A).

Materials
Participants received a printed Brief (see Appendix B) and Consent Form (see Appendix C). A measuring tape and chin strap was used to secure the wireless electrode cap. Sponges, saline solution, hypoallergenic gel and clean gauze were used for conductance of electrodes. Electrode location followed the international 10-20 EEG system (Jasper, 1958), using Fp1, Fp2, Pz, M1, M2 (see Figure 1). Also, electrodes DRL and CMS were used to allow the SMARTING system (Mobile SMARTING 24-channel sleep EEG amplifier, mBrainTrain, Belgrade, Serbia) to work. For the task, a 21-inch HP monitor with a screen resolution of 1920x1080 pixels and a refresh rate of 100Hz showed the instructions, fixation dot and stimuli. Participants sat in a chair 115cm from the screen. The stimuli consisted of five random dates and the subjects' date of birth, all 2x6cm corresponding to 1x3º of visual angle, and in written format (see Figure 2). Participant responses were made on a Cedrus box. The amplifier recorded with a resolution of 24 bits and a sampling rate of 250Hz and improved the signal strength. Bluetooth was used to transmit data from the amplifier to a smartphone and SMARTING Windows. A Lab Streaming Layer framework (GitHub, 2021) was used to combine the EEG and task data. Saline solution was used to clean the sponges. All participants received a written Debrief (see Appendix D).

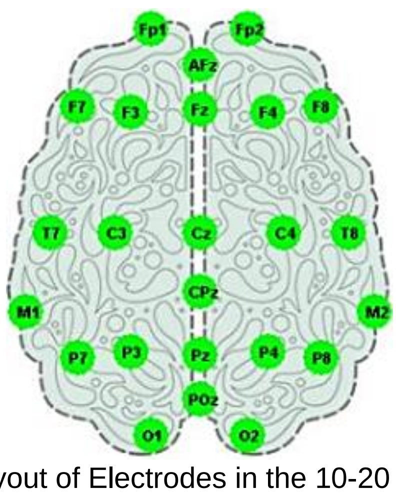

Figure 1: Layout of Electrodes in the 10-20 EEG System.

Figure 2: Example of a Date Presented.

Procedure

Upon arrival at the lab, participants were given a Brief (see Appendix B) to read and a Consent Form (see Appendix C) to sign. Their head circumference was measured using measuring tape to ensure the cap was fitted tightly. The wireless EEG cap was then placed on their head and secured using the chin strap. Participants' hair was parted where the electrodes were going to be placed to ensure their brainwaves could be detected. Sponges were soaked in saline solution for conductance and placed on the end of 7 electrodes. Fp1 and Fp2 were used to measure movement artifacts by detecting eye movements and blinking. The electrodes were attached to the electrode cap in accordance with the 10-20 EEG system (see Figure 1). An amplifier was secured onto the back of the cap to increase the strength of the signal. The connectivity of the electrodes to the SMARTING programme was then tested. If the electrodes were out of range, hypoallergenic gel was applied to increase conductance. A clean gauze was used to remove any excess. This set-up took approximately 5-10 minutes for each participant.

Once connectivity was ensured, participants were taken to an adjoining dark room to limit distractions whilst completing the experiment. They sat in a comfortable chair and were given a Cedrus box with yes/no options. The participants were then told to relax and attempt to minimise eye movements and blinking in order to limit movement artifacts. They were informed of their target date, which was a random date of no personal significance. The instructions appeared on-screen telling the participants to focus on the fixation spot at the centre of the screen which would precede the presentation of the stimuli. They were informed that for each stimuli presented a button press of yes or no was required to indicate their recognition. They were told to try to conceal their date of birth by responding with a button press of 'yes' to their target date, and 'no' to all other stimuli. To check their understanding, they were asked to verbally summarise the instructions.

The participant was then left alone in the room to complete the practice trial. This lasted for 1 minute and consisted of 15 stimulus presentations (see Figure 2). After this, the electrode functioning and participants' responding to target stimuli was verified. They then completed the experiment which consisted of two 10-minute blocks of 210 trials each, with 70 presentations per item type. They had the option of a short break between the two blocks to decrease fatigue. Each stimulus flashed on the screen for 200ms, with an average interstimulus interval of 2000ms (1900-2100ms), followed by the participant's button press with their dominant hand to ensure their attention. The stimuli were presented in a random order with the probe stimuli (their date of birth) appearing 20% of the time. Overall, the study lasted approximately 20 minutes per participant.

The data (continuous brain waves from each electrode) was transmitted via Bluetooth to the smartphone and was recorded on SMARTING Windows. It used online analog-digital conversion to allow for the signals to be processed on the computer. This EEG data was then combined with the task information (the button presses and stimuli presentation times) from the other monitor using a Lab Streaming Layer Framework. This created an '.xdf-file' which showed the ERPs (the brainwave response to stimuli presentation) and the participant reaction times (the time taken to respond via button press). These files were stored on a disk for analysis later.

After completion of the study, the electrodes were taken out and the sponges were cleaned in saline solution. The electrode cap was then removed. Each participant was given a Debrief (see Appendix D) to take away with them. This explained the experiment and the underlying theory and provided a point of contact if they had any further questions or wished to withdraw their data.

EEG Analysis

In accordance with existing literature, the data was re-referenced off-line to the average of the two mastoids. Off-line, ERPs were averaged for an epoch of 1500ms (including a 200ms baseline). Contamination was evident within some trials, due to eye movements or blinking. Where possible, these were cleaned using the Independent Component Analysis (Makeig et al., 1996). For some trials there were remaining artifacts (less than 1%). These were rejected off-line.

To test the minimum number of trials required to accurately detect deception, the data was split into subsections of 10,20,30,40,50 and 60 trials per stimulus. The highest was 60, as opposed to 70, due to the discarding of some data due to movement artifacts.

Also, a sample of virtual innocent participants was created by declaring a random irrelevant as the probe and comparing them to the other irrelevants. The statistical tests on guilty participants compared the probe to 3 out of 4 of the irrelevants in order to ensure accurate comparisons between the guilty and innocent datasets.

Results

Bootstrap Technique

The bootstrap technique was used to test the hypothesis that the accuracy of deception detection increases with the number of trials. The average of the probes/virtual probes and irrelevants was calculated for the guilty and innocent dataset for each subject and number of trials. Data points were then randomly assigned as probes or irrelevants and the means were calculated. This was repeated for 1000 permutations (to create a distribution of values) and compared to the actual data. Those below the $p=0.1$ threshold level of significance (as recommended by Rosenfeld, 2011) were correctly classified as guilty, and those above as innocent. This was also tested against the thresholds $p=0.5/0.25/0.5$. For each number of trials, the mean number of correctly identified guilty/innocent participants was calculated and converted into a percentage which is recorded as lines in Figure 3. It is evident from Figure 3 that the threshold levels of significance of $p=0.5$ and $p=0.25$ cannot be used as the lines lie further from the average, whereas the standard threshold of p=0.1 is closer. The threshold of p=0.15 shows a similar pattern of results as p=0.1. However, it reveals higher accuracy at 40 trials, suggesting that it may be a better threshold.

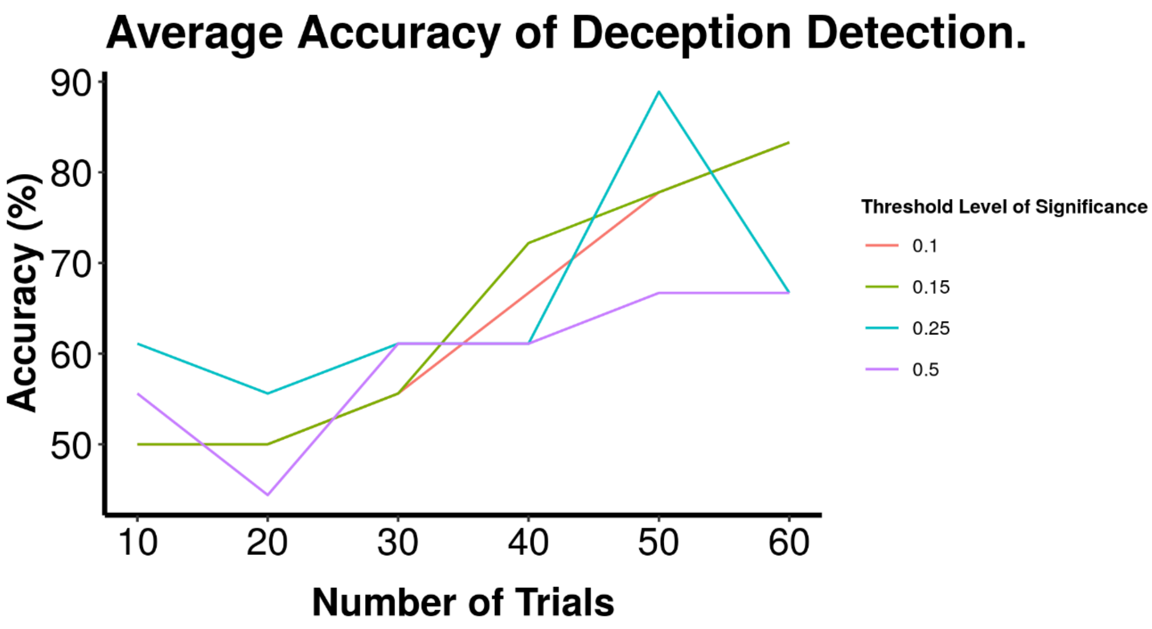

Figure 3: Accuracy of Deception Detection at each Number of Trials.

The line representing the threshold p=0.15 shows that at 10 trials there is an average of 50% accuracy, meaning half the participants were undetected. Therefore, accuracy at 10 trials was by chance. Whereas, at 60 trials there is an average of approximately 85% accuracy, revealing that the majority of participants were correctly detected. These results indicate a positive relationship between accuracy and number of trials. Thus, Figure 3 suggests that the higher the number of trials the more accurate the results for single-person identification of guilt/innocence.
The threshold of p=0.15 shows that 10 to 30 trials should not be used as the accuracy is less than 60%. The increase in accuracy flattens from 40 to 60 trials and peaks at 60 trials suggesting that this is the best number of trials to use. However, conducting 50 trials would be less time-consuming and costly and only slightly less accurate (by approximately 5%), which suggests that in certain situations this could be more beneficial.

Bayesian Anova

A Bayesian Anova was used to test for an effect of stimuli type (probe/irrelevant) and the number of trials (10 to 60) on amplitude, and the interaction between the two. This was a within-subject design as the guilty dataset was used to create virtual innocent participants.

The result for an effect of the number of trials used is less than 0.33 (BF=0.039) and therefore non-significant. This finding provides evidence for the null hypothesis, revealing an absence of effect of number of trials on amplitude. Although Figure 3 indicates that more trials increased accuracy at an individual level, due to less noise at 60 trials than 10 (see Figure 5). Whereas, the Anova tested at a group level where the average washed out the noise.

However, a main effect of stimuli type was found (BF=2.43X10^{17}), which provides significant evidence for the experimental hypothesis as opposed to the null. Therefore, it supports the hypothesis that the item type (probe/irrelevant) effects the amplitude. When testing for an interaction between stimuli type and number of trials, the result are non-significant (BF=0.105). This reveals substantial evidence for an absence of interaction, highlighting that the difference in responses to stimuli type does not vary by the number of trials.

Direction of effect of item type
Figure 4 was created to show the direction of the effect of item type found in the Anova (BF=2.43X10^{17}). For each number of trials, the mean difference in amplitude between probes and irrelevants for all subjects combined was calculated. This was done for both the guilty and innocent datasets. The error bars show the standard deviations.

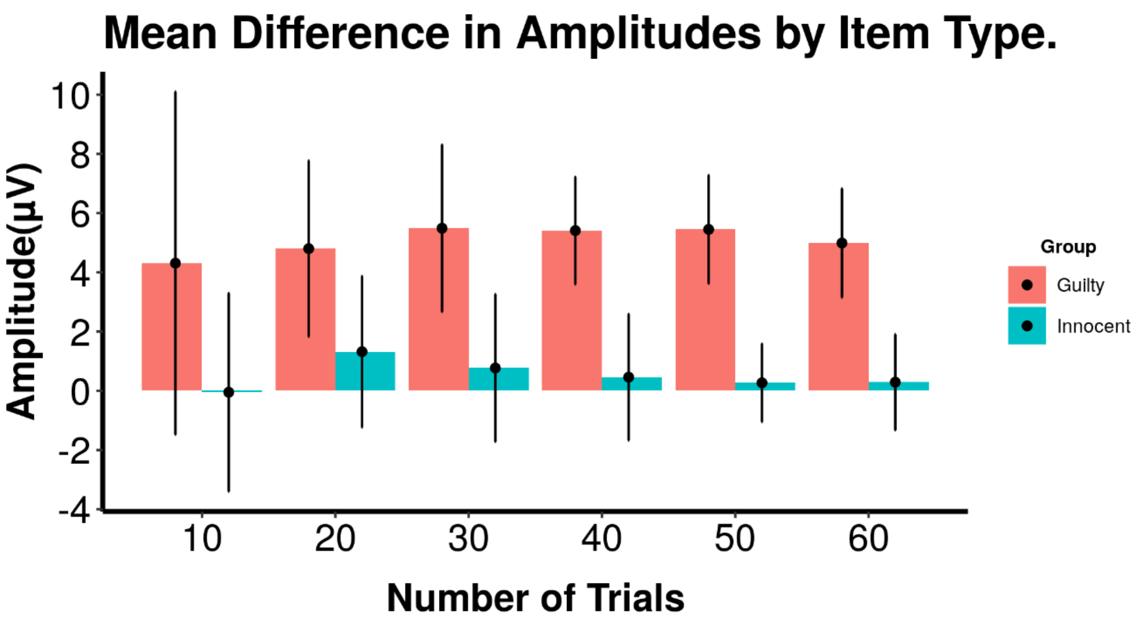

Figure 4: *Differences in Amplitudes at each Number of Trials.*

Figure 4 reveals that guilty participants consistently had a bigger difference in amplitude between probe and irrelevant stimuli than the virtual innocent participants. This suggests that the guilty participants had a bigger brainwave response to probe items than innocent participants, highlighting the accuracy of the EEG in detecting the P300 response to deception.

The standard deviations shown by the error bars in Figure 4 reveal that the lower the number of trials (10 to 30), the greater the variability in the data. This could be due to increased noise at 10 trials as opposed to 60 (see Figures 5&6). This reiterates the idea that 10 to 30 trials should not be used as the margin for error is too large.

Line Graphs

Figures 5 and 6 show the brainwave responses to each stimuli type. Each line represents the mean amplitude of all participants. Figure 5 shows that at 10 trials, there is higher trial to trial variability due to more noise. This can be seen by the overlapping lines and the response to the probe being harder to detect. This shows that conducting 10 trials was not enough to accurately detect deception. Whereas Figure 6 reveals that at 60 trials there is a clear P300 response to the probe. This reveals that with increasing number of trials, the clarity of the P300 response improves, suggesting 60 trials to be optimum.

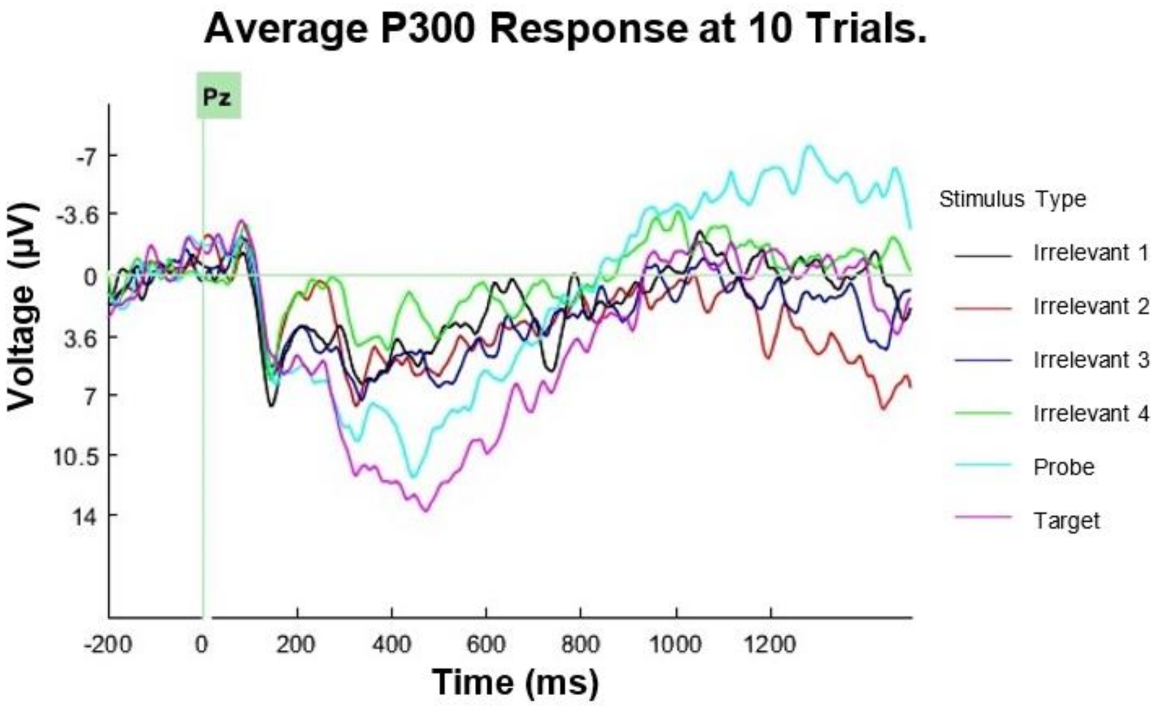

Figure 5: *P300 Responses to each Stimuli Type at 10 Trials.*

The conclusion that at a group level the number of trials effects the accuracy of results, differs from that of the Bayes Factor. Figure 6 reveals that the peaks and troughs of wavelengths per item type cancel out at 60 trials. However, the combination of subject’s brainwave responses is not enough to stop the fluctuations per item type at 10 trials. Whereas, in the Bayes Factor, the group average washes out the noise, eliminating the effect of number of trials.

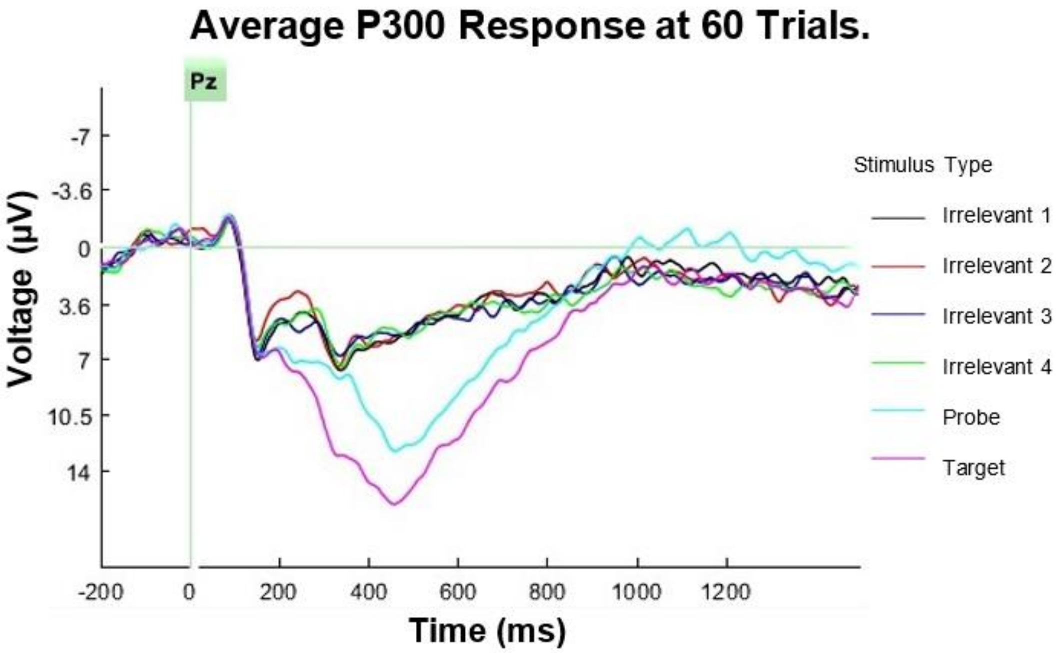

Figure 6: *P300 Responses to each Stimuli Type at 60 Trials.*

Means and Standard Deviations

Table 1 shows the means and standard deviations of amplitude for each item type per number of trials. The means reveal that the probes consistently produce higher amplitude brainwave responses than the irrelevant. Also, there is higher variability in amplitude responses with smaller number of trials. This is evident by the standard deviation for probe stimuli at 10 trials (SD=8.46) being larger than at 60 (SD=2.79).

Table 1: *Means and Standard Deviations per Stimuli Type per Number of Trials.*

Trials	Probe		Target		Irrelevants	
	Mean	SD	Mean	SD	Mean	SD
10	8.71	8.46	11.74	8.27	4.39	6.46
20	9.54	5.35	12.50	5.48	5.07	4.56
30	10.45	4.33	12.98	4.57	5.15	3.76
40	10.54	3.48	13.31	3.68	5.25	3.19
50	10.71	3.32	13.63	3.66	5.43	2.89
60	10.42	2.79	13.92	3.60	5.50	2.72

Discussion

This study demonstrated the most appropriate number of trials from which concealed information can be accurately detected using an EEG. This technique was proven to accurately detect concealed information, as participants had significantly different brainwave responses to each item type (probes and irrelevants). Verschuere et al. (2011) described the recognition of stimuli to be evident through an ERP peak (the P300) 300-800ms after the event. This study confirms this as Figure 6 highlights participants' recognition of the probe stimuli, as 450ms after the event there was a P300 response. Figure 6 shows that the P300 response is also elicited by the target stimuli, this peak was greater than the response to the probes. Rosenfeld (1998) explained this effect as the response task encourages participants to attend to the target stimuli. Also, Solokov (1963) found an increased orientating response when stimuli were significant to the subject. This effect is replicated in this study as the Bayesian Anova reveals that the item type has a significant effect on the amplitude of brainwave responses. This indicates that the probes, which were of personal significance, caused a greater P300 response than the irrelevants which were insignificant. This is supported by the findings of Johnson (1986) who found that the probes elicited a larger P300 response as they were more meaningful. Figure 4 highlights that guilty participants' responses had a bigger difference in amplitude, indicating a greater response to probe stimuli than innocent participants'. Therefore, this study supports previous research in confirming that brainwave responses to probe stimuli can reveal concealed information.

Furthermore, this experiment tested the accuracy of detecting deception using an EEG. Figure 3 reveals that the EEG has 85% accuracy in classifying participants as guilty or innocent (using the p=0.15 threshold level of significance). This contrasts with Lykken (1959) who found 93.9% correct classification of guilt using the CIT. However, it supports Rosenfeld (2011) who used an EEG and found 85-100% accuracy due to meaningful stimuli producing a greater P300 response. The findings imply that using an EEG is a reliable method to accurately detect concealed information, although there was evidence of countermeasures. Ganis et al. (2011) questioned the utility of using fMRI due to countermeasures decreasing the accuracy of results. However, despite attempts to limit countermeasures such as movements, they were still evident within this EEG study. Therefore, some data was discarded and up to 60 trials was used instead. Despite this, the results show that the EEG is highly accurate in detecting concealed information.

This study also investigated the effects of using different numbers of trials. Figures 5 and 6 confirm the findings of Scheer et al. (2005) as they show decreased clarity of results due to noise. This is seen by the fluctuation and overlapping of the lines at 10 trials, making the brainwave responses to different item types less clear. Ganis and Schendan (2013) found that increasing the number of trials reduced the signal-to-noise ratio. This study supports this, as seen by Figure 4 which reveals less variability in higher number of trials, shown by the error bars representing standard deviations in amplitude. Therefore, the findings suggest that more trials would be beneficial as it would ensure clearer results due to decreased effects of noise. However, Lynn (2013) argued that increasing the number of repeated presentations of stimuli causes the orientation effect to habituate, resulting in no response to stimuli. This study challenges this, as Figures 5 and 6 show that the clarity of the P300 response increases with the number of trials, suggesting that repeated presentations do not cause habituation of the orientation effect.

This experiment revealed that the EEG is a reliable detector of concealed information. This study supports the findings of Lefebvre et al. (2007) who found the P300 response to be a reliable indicator of recognition. This is seen in Table 1, where the means of the probe stimuli are consistently higher than the means of the irrelevant stimuli across each number of trials. This highlights the reliability of this method as different responses to each item type were always found. However, this study also expanded on research by Lefebvre et al. (2007) as it found that the reliability of this method is affected by the number of trials. This is evident in Table 1, which shows that the smaller the number of trials the greater the variability in amplitudes. This implies that accurate detection would be harder to achieve at 10 trials than at 60 trials. The Anova results reveal no effect of number of trials, however this was conducted at a group level, whereas in real life this technique would be applied on an individual level (which the results show is affected by number of trials). Figure 3 shows the accuracy of detection increasing until 40 trials and peaking at 60 trials. However, due to 50 trials not being significantly less accurate than 60 and less time-consuming and costly, it may be the most appropriate number of trials to use, depending on context. This would have positive implications for the issue outlined by Rosenfeld (2011), whereby forensic agencies are reluctant to run an EEG due to the effort and expense required to learn and use this technique, as it would make this method easier to conduct.

Limitations

Due to the Coronavirus precautions this study was limited in its sample size (9 participants), thus making the results of limited utility. Furthermore, the participants were tested immediately, whereas in the real-world EEG testing is likely to occur further from the time of the event. This may alter the accuracy of this measure. Also, if participants were not of a sound mind (for example, suffering from psychosis) they may believe their lies, thus inhibiting a brainwave response. Hence, the accuracy of this method may well be context specific.

A limitation of using the CIT and EEG method is that it may be difficult for non-specialists to think of enough irrelevant stimuli (to compare to probes) and to assemble a scientifically valid CIT, particularly when cases are non-specific, such as for employee screening as opposed to criminal investigations. Therefore, despite this study suggesting relatively limited time is required to conduct the experiment, there may still be issues with the time taken to create the CIT. The existence of countermeasures may exaggerate this, as more trials would need to be conducted to allow for the discarding of unreliable data.

This study revealed deception detection to be accurate as it was conducted in a controlled environment. However, in criminal investigations, innocent participants may produce false positives through non-crime related memories. Also, guilty participants could produce false negatives due to memory distortions of details of the crime. Furthermore, innocent participants may be exposed to critical material (for example, through the media) prior to testing which would limit the validity of results.

Conclusions

Overall, this study confirmed the accuracy of using CIT and EEG to detect deception and determined 50 as the recommended number of trials. This allows for high accuracy and clear P300 results, which are used to reveal recognition of information. This research was of high salience as it has significant implications in law enforcement (such as criminal investigations) and occupations (for example employee screening and security clearances). By outlining the minimum number of trials required, it makes this method less laborious and time-consuming, thus facilitating wider use. Therefore, by more people adopting this technique, it would increase the amount of deception detected, which has the potential to reduce serious negative consequences (such as crimes) for both individuals and society as a whole.

Future Work

To expand on this study, a replication using the Complex Trial Protocol (CTP) could be conducted to test the minimum number of trials needed to detect deception in this method compared to the CIT. The results may differ using the CTP, as this test is of higher accuracy and more resistant to countermeasures. Rosenfeld et al. (2008) found up to 90% accuracy using a CTP and discovered that the P300 was less vulnerable to interference from countermeasures, contrasting with the CIT. Therefore, the CTP may be a more reliable method. Although, as it is a more complex procedure, it may still be of lower utility than the CIT as it may take longer to run despite using the minimum number of trials required.

Additionally, to test the validity of the results, future studies could incorporate a delay between learning about the probe stimuli and the CIT, as this is more representative of how this test would actually be used. This study had no delay and was conducted in lab settings, whereas the proposed study would test the accuracy of this measure in conditions closer to that of the real world. This may decrease the accuracy of the test as it allows more time for memory distortions, source confusion and rehearsal of information. Loftus and Hoffman (1989) found that information after the event can distort memories and lead to the creation of new ones. Therefore, it is important to test the accuracy of this measure when subjected to real world extraneous variables.

Acknowledgements

I would like to thank my supervisor, Dr Giorgio Ganis, for his guidance and enthusiasm throughout this study. As always, I would like to express my gratitude to my friends and family for their continued and unwavering support.

References

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development*, *63*(1), 129-137.

Carson, T. L. (2010). *Lying and deception: Theory and practice*. Oxford University Press.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology*, *70*(5), 979.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in psychophysiology*, *2*(S 1), 78.

Ganis, G., & Schendan, H. E. (2013). Concealed semantic and episodic autobiographical memory electrified. *Frontiers in human neuroscience*, *6*, 354.

Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., & Schendan, H. E. (2011). Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage*, *55*(1), 312-319.

GitHub (2021). Lab Streaming Layer. Retrieved from https://github.com/sccn/labstreaminglayer

Grazioli, S. (1998). Detecting manipulations of financial information.

Hartwig, M., Anders Granhag, P., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, *13*(2), 213-227.

Honts, C. R., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, *79*(2), 252.

Jasper, H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr Clin Neurophysiol*, *10*, 370-375. doi:10.1016/0013-4694(58)90053-1

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*.

King, W. R., & Dunn, T. M. (2010). Detecting deception in field settings. *Policing: An International Journal of Police Strategies & Management*.

Klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2016). Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures. *Psychophysiology*, *53*(4), 579-590.

Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology*, *44*(6), 894-904.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of experimental psychology: General*, *118*(1), 100.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385.

Lynn, R. (2013). *Attention, Arousal and the Orientation Reaction: International Series of Monographs in Experimental Psychology* (Vol. 3). Elsevier.

Makeig, S., Jung, T. P., Ghahremani, D., & Sejnowski, T. J. (1996). Independent component analysis of simulated ERP data. *Institute for Neural Computation, University of California: technical report INC-9606*.

Masip, J., Garrido, E., & Herrero, C. (2004). Defining deception. *Anales de psicologia*.

Misaka, R., Hira, S., & Furumitsu, I. (2009). Comparison of auditory and visual stimulus presentation during the P300-based concealed information test. Psychophysiology, *46*, 131.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. *Memory detection: Theory and application of the Concealed Information Test*, 63-89.

Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology*, *45*(6), 906-919.

Rosenfeld, J. P., Reinhart, A. M., Bhatt, M., Ellwanger, J., Gora, K., Sekera, M., & Sweet, J. (1998). P300 correlates of simulated malingered amnesia in a matching-to-sample task: Topographic analyses of deception versus truthtelling responses. *International Journal of Psychophysiology*, *28*(3), 233-247.

Scheer, H. J., Sander, T., & Trahms, L. (2005). The influence of amplifier, interface and biological noise on signal quality in high-resolution EEG recordings. *Physiological measurement*, *27*(2), 109.

Sokolov, E. N. (1963). Perception and the conditioned reflex.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Punishment: Issues and experiments*, 27-41.

Appendices are provided separately as supplementary files (see additional downloads for this article).